

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 2001, 18:24:52 ; Search time 113.17 Seconds  
(without alignments)  
225.225 Million cell updates/sec

Title: US-09-373-230-5  
Perfect score: 17  
Sequence: 1 TTYGARGARATGAYCC 17

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1283235 segs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_bal:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_bal:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_rod:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_vi:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_vil:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_ro1:\*  
95: gb\_ro2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.4	90.6	17	9 AR072047	AR072047 Sequence
2	15.4	90.6	471	9 AR072044	AR072044 Sequence
3	15.4	90.6	471	10 E13264	E13264 Mouse cDNA
4	15.4	90.6	471	10 E14257	E14257 CDNA encodi
5	15.4	90.6	471	10 E14760	E14760 CDNA encodi
6	15.4	90.6	471	10 E17139	E17139 Murine mRNA
7	15.4	90.6	471	45 E10609	E10609 Mouse CDNA
8	15.4	90.6	471	45 E11744	E11744 CDNA encodi
9	15.4	90.6	471	45 E12010	E12010 CDNA encodi
10	15.4	90.6	471	56 ASIGIF	Y09278 Artificial
11	15.4	90.6	572	94 MMU66244	U66244 Mus musculu

12 15.4 90.6 866 94 MUSIGIFPP D49949 Mouse mRNA  
13 15.4 90.6 35143 90 AP000229 AP000229 Homo sapi  
14 15.4 90.6 59955 77 AC080029 AC080029 Homo sapi  
15 15.4 90.6 95155 88 AC078950 AC078950 Homo sapi  
16 15.4 90.6 100000 90 AP000144 AP000144 Homo sapi  
17 15.4 90.6 100634 90 AP001594 AP001594 Homo sapi  
18 15.4 90.6 123631 91 HS22F01 AL109967 Homo sapi  
19 15.4 90.6 202103 87 AC010176 AC010176 Homo sapi  
20 15.4 90.6 340000 90 AP001695 AP001695 Homo sapi  
21 14.4 84.7 355 54 G01707 G01707 chicken STS  
22 14.4 84.7 710 94 MMHBEFL4 U39192 Mus musculu  
23 14.4 84.7 963 53 CNS01G2U AL143803 Anopheles  
24 14.4 84.7 997 54 CNS07CNO AL439404 T7 end of  
25 14.4 84.7 1084 53 CNS07OT9 AL424051 clone AZO  
26 14.4 84.7 1188 5 AF318610 AF318610 Caenorhab  
27 14.4 84.7 1718 12 AF046922 AF046922 Colletotr  
28 14.4 84.7 1971 4 AB035057 AB035057 Drosophil  
29 14.4 84.7 2094 4 AB035066 AB035066 Drosophil  
30 14.4 84.7 2178 94 AB035069 AB035069 Drosophil  
31 14.4 84.7 2237 4 AB035062 AB035062 Drosophil  
32 14.4 84.7 2268 4 AB035061 AB035061 Drosophil  
33 14.4 84.7 2282 4 AB035061 AB035061 Drosophil  
34 14.4 84.7 2305 4 AB035065 AB035065 Drosophil  
35 14.4 84.7 2919 12 AF024633 AF024633 Magnapor  
36 14.4 84.7 3172 12 AF015753 AF015753 Magnapor  
37 14.4 84.7 3522 4 AB035058 AB035058 Drosophil  
38 14.4 84.7 4438 8 AF085184 AF085184 Gallus ga  
39 14.4 84.7 13952 14 R1C1CPA D32136 Rice endoge  
40 14.4 84.7 20128 13 AF155773 AF155773 Gibberell  
41 14.4 84.7 21077 5 CEK10C8 Z74474 Caenorhabdi  
42 14.4 84.7 21644 92 HSH2A1KLT Z81330 Human DNA s  
43 14.4 84.7 22092 4 AC006637 AC006637 Caenorhab  
44 14.4 84.7 24361 5 CELR02C2 AF039037 Caenorhab  
45 14.4 84.7 27748 5 CELH34C03 AF100662 Caenorhab

## ALIGNMENTS

RESULT 1  
AR072047 17 bp DNA PAT 18-FEB-2000  
LOCUS AR072047  
DEFINITION Sequence 5 from patent US 5912324.  
ACCESSION AR072047  
VERSION AR072047.1 GI:72222935  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 17)  
AUTHORS Okamura,H., Tanimoto,T., Torigoe,K., Kunikata,T., Taniguchi,M.,  
Kohno,K. and Kurimoto,M.  
TITLE Interferon-gamma (IFN-gamma.) inducing factor (IGIF, IL-18)  
JOURNAL purified from murine liver  
FEATURES Patent: US 5912324-A 5 15-JUN-1999;  
source Location/Qualifiers  
1..17  
BASE COUNT 4 a /organism="unknown"  
ORIGIN 2 c 4 g 3 t 4 others

Query Match 90.6%; Score 15.4; DB 9; Length 17;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAYCC 17  
Db 1 TTYGARGARATGAYCC 17

RESULT 2  
AR072044

LOCUS AR072044 471 bp DNA PAT 18-FEB-2000  
DEFINITION Sequence 1 from patent US 5912324.  
ACCESSION AR072044  
VERSION AR072044.1 GI:72222932  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 471)  
AUTHORS Okamura,H., Tanimoto,T., Torigoe,K., Kunikata,T., Taniguchi,M.,  
Kohno,K. and Kurimoto,M.  
TITLE Interferon-gamma (IFN-gamma.) inducing factor (IGIF, IL-18)  
JOURNAL purified from murine liver  
FEATURES Patent: US 5912324-A 1 15-JUN-1999;  
source Location/Qualifiers  
1..471  
BASE COUNT 162 a /organism="unknown"  
ORIGIN 91 c 92 g 125 t 1 others

Query Match 90.6%; Score 15.4; DB 9; Length 471;  
Best Local Similarity 76.5%; Pred. No. 4.5e+02;  
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAYCC 17  
Db 244 TTTGAGGAATGATCC 260

RESULT 3  
E13264 471 bp DNA PAT 24-JUN-1998  
LOCUS E13264  
DEFINITION Mouse cDNA encoding a protein that induces to produce  
interferon-gamma.  
ACCESSION E13264  
VERSION E13264.1 GI:3252069  
KEYWORDS JP 1997157180-A/2.  
SOURCE Mus sp.  
ORGANISM Mus sp.  
REFERENCE 1 (bases 1 to 471)  
AUTHORS Torigoe,K., Tanimoto,T., Fukuda,S. and Kurimoto,M.  
TITLE AGENT FOR SENSITIVE DISEASE  
JOURNAL Patent: JP 1997157180-A 2 17-JUN-1997;  
HAYASHIBARA BIOCHEM LAB INC  
COMMENT OS Mus sp. (mouse)  
PN JP 1997157180-A/2  
PD 17-JUN-1997  
PE 24-JAN-1996 JP 1996028722  
PR 10-MAR-1995 JP 95P 78357, 29-SEP-1995 JP 95P 274988, PR  
04-OCT-1995 JP 95P 279906  
PI TORIGOE KAKUJI, TANIMOTO TADAO, FUKUDA SHIGETSU, PI  
KURIMOTO MASASHI  
PC A61K38/00,A61K38/00,A61K38/00,A61K38/00,A61K38/00,C07K14/52,  
PC C07K14/54,  
PC C07K14/55;  
CC strandedness: Double;  
CC topology: Linear;  
CC Feature is identified by similarity;  
FH Key Location/Qualifiers  
FT source 1..471  
FT /organism="Mus sp."  
FT /tissue\_type="liver"  
FT mat\_peptide 1..471  
FT /product="interferon-gamma inducer protein".  
FT location/Qualifiers  
1..471  
FT /organism="Mus sp."  
FT /db\_xref="taxon:10095"

BASE COUNT 162 a 91 c 92 g 125 t 1 others



FEH	Key	Location/Qualifiers
FT	source	1. .471
FT		/organism='Mus sp. '
FT		/tissue_type='liver'
FT	mat_peptide	1. .471
FT		/product='IL-18'.
FEH	source	1. .471
BASE COUNT	162 a	91 c 92 g 126 t
ORIGIN		
		/organism="Mus sp. "
		/db_xref="taxon:10095"
Query Match	90.6%;	Score 15.4; DB 10; Length 471;
Best Local Similarity	76.5%;	Pred. No. 4.5e+02;
Matches 13; Conservative	4;	Mismatches 0; Indels 0; Gaps 0;
QY	1 TTYGARGARATGGAYCC 17	
	: : : : : : :	
Db	244 TTTGAGGAATGATGCC 260	
RESULT	7	
E10609		
ID	E10609	standard; RNA; ROD; 471 BP.
XX	AC	E10609;
XX	SV	E10609.1
XX		
DT	08-OCT-1997 (Rel. 52, Created)	
DT	02-SEP-2000 (Rel. 65, Last updated, Version 2)	
DE	Mouse cDNA encoding a protein involved in interferon-gamma production.	
XX		
KW	JP 1996027189-A/1.	
XX		
OS	Mus sp.	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;	
OC	Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.	
XX		
RN	[1]	
RP	1-471	
RA	Okamura H., Tanimoto T., Torigoe K., Kurimoto M.;	
RT	"PROTEIN INDUCING PRODUCTION OF INTERFERON-GAMMA";	
RL	Patent number JP1996027189-A/1, 30-JAN-1996.	
XX	HAYASHIBARA BIOCHEM LAB INC.	
OS	Mus sp. (mouse)	
PN	JP 1996027189-A/1	
CC	PD 30-JAN-1996	
CC	PF 14-JUL-1994 JP 1994184162	
CC	PI OKAMURA HARUKI, TANIMOTO TADAO, TORIGOE KAKUJI,	
CC	PI KURIMOTO MASASHI	
CC	PC C07K14/52, A61K38/00, A61K38/00, C12N1/21, C12N15/09,	
CC	C12P21/02//C07K14/57;	
CC	PC C07K14/52, A61K38/00, A61K38/00, C12N1/21, C12N15/09,	
CC	CC strandedness: Double;	
CC	CC topology: Linear;	
CC	CC key	
CC	CC FH	Location/Qualifiers
CC	CC FT	1. .471
CC	CC FT	/organism="Mus sp. "
CC	CC FT	/tissue_type="liver"
CC	CC FT	1. .471
CC	CC FT	/product="a protein involved in
CC	CC FT	interferon-gamma
CC	CC FT	production"
XX		
Key	Location/Qualifiers	
FT	source	1. .471

```

FT      /db_xref="taxon:10095"
FT      /organism="Mus sp."
XX
SQ      Sequence 471 BP; 162 A; 91 C; 92 G; 125 T; 1 other;

Query Match          90.6%; Score 15.4; DB 45; Length 471;
Best Local Similarity 76.5%; Pred. No. 4.5e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTYGARGARATGAYCC 17
        ||:||:||:||:||
Db      244 TTTGAGGAAATGATCC 260

RESULT      8
E11744
ID      E11744      standard; RNA; ROD; 471 BP.
XX
XX      E11744;
AC
XX      E11744.1
SV
XX      08-OCT-1997 (Rel. 52, Created)
DT      02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX
XX      cDNA encoding polypeptide which induce mouse interferon-gamma product.
DE
XX      JP 1996193098-A/1.
KW
XX      Mus musculus (house mouse)
OS
XX      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX
XX      [1]
RN      1-471
RP      Ushio S., Torigoe K., Tanimoto T., Okamura H., Kurimoto M.;
RA      "POLYPEPTIDE FOR INDUCING PRODUCTION OF INTERFERON-GAMMA";
RT      Patent number JP1996193098-A/1, 30-JUL-1996.
RL      HAYASHIBARA BIOCHEM LAB INC.
XX
XX      OS Mus musculus (mouse)
CC      PN JP 1996193098-A/1
CC      PD 30-JUL-1996
CC      PF 18-SEP-1995 JP 1995262062
CC      PR 15-NOV-1994 JP 94P 304203
CC      PI USHIO SHINPEI, TORIGOE KAKUJI, TANIMOTO TADAO, OKAMURA HARUKI,
CC      PI KURIMOTO MASASHI
CC      PC C07K14/52,C07H21/04,C12N1/21,C12N15/09,C12P21/00//A61K38/00,
CC      PC C07K7/06,
CC      PC C07K7/08,(C12N1/21,C12R1:19),(C12P21/00,C12R1:19);
CC      CC strandedness: Double;
CC      CC topology: Linear;
CC      CC hypothetical: No;
CC      CC anti-sense: No;
CC      CC FH Key
CC      CC FH Location/Qualifiers
CC      CC FH
CC      CC FT source 1..471
CC      CC FT /db_xref="taxon:10090"
CC      CC FT /organism="Mus musculus"
CC      CC FT /tissue_type="liver"
XX
XX      Key Location/Qualifiers
XX      FT source 1..471
XX      FT /db_xref="taxon:10090"
XX      FT /organism="Mus musculus"
SQ      Sequence 471 BP; 162 A; 91 C; 92 G; 125 T; 1 other;

Query Match          90.6%; Score 15.4; DB 45; Length 471;
Best Local Similarity 76.5%; Pred. No. 4.5e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TTYGARGARATGAYCC 17  
||:||||:||||:|  
Db 244 TTTGAGGAATGATGCC 260

RESULT 9  
ID E12010 standard; RNA; ROD; 471 BP.  
XX E12010;  
XX E12010.1  
SV E12010.1  
DT 07-OCT-1997 (Rel. 52, Created)  
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)  
XX  
DE cDNA encoding mouse polypeptide which introduce interferon-gamma product in  
DE immunocompetent cell.  
XX JP 1996231598-A/2.  
KW  
XX Mus musculus (house mouse)  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
XX  
RN [1]  
RP Kunikata T., Taniguchi M., Kono K., Kurimoto M.;  
RA "MONOCLONAL ANTIBODY";  
RT Patent number JP1996231598-A/2, 10-SEP-1996.  
RL HAYASHIBARA BIOCHEM LAB INC.  
XX  
XX OS Mus musculus (mouse)  
CC PN JP 1996231598-A/2  
CC PD 10-SEP-1996  
CC PF 23-FEB-1995 JP 1995058240  
CC PI KUNIKATA TOSHIO, TANIGUCHI MUTSUOKO, KONO KEIZO,  
CC PI KURIMOTO MASASHI  
CC PC C07K16/24, C07K1/16, C07K1/18, C07K1/22, C07K1/26, C07K1/30,  
CC PC C07K1/34, C12N5/10,  
CC PC C12N15/02, C12P21/08, G01N33/53, G01N33/577//A61K38/21, A61K39/395,  
CC PC (C12P21/08,  
CC PC C12R1:91);  
CC CC strandedness: Double;  
CC CC topology: linear;  
CC CC hypothetical: No;  
CC CC anti-sense: No;  
CC FH key Location/Qualifiers  
CC FH source 1..471  
CC FT /organism="Mus musculus"  
CC FT /tissue\_type="liver"  
CC XX  
FH key Location/Qualifiers  
FH FH source 1..471  
FH FT /db\_xref="taxon:10090"  
FH FT /organism="Mus musculus"  
XX  
SQ Sequence 471 BP; 162 A; 91 C; 92 G; 125 T; 1 other;

Query Match 90.6%; Score 15.4; DB 45; Length 471;  
Best Local Similarity 76.5%; Pred. No. 4.5e+02;  
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYGARGARATGAYCC 17  
||:||||:||||:|  
Db 244 TTTGAGGAATGATGCC 260

RESULT 10

ASIGIF 471 bp DNA SYN 08-NOV-1996  
LOCUS Artificial sequence DNA for cytokine interferon-gamma inducing  
DEFINITION factor (IGIF).  
ACCESSION Y09278  
VERSION Y09278.1 GI:1666283  
KEYWORDS cytokine; interferon-gamma inducing factor (IGIF).  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 471)  
AUTHORS Schluesener, H.J.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 471)  
AUTHORS Schluesener, H.J.  
TITLE Direct Submission  
JOURNAL Submitted (21-OCT-1996) H.J. Schluesener, Inst.f. Hirnforschung,  
Calwer Str. 3, D-77076 Tuebingen, FRG

FEATURES  
source  
1..471  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
CDS  
1..>471  
/codon\_start=1  
/transl\_table=1  
/product="mouse cytokine interferon-gamma inducing factor"  
/protein\_id="CAA70473.1"  
/db\_xref="GI:1666284"  
/translation="NFGRLHCTTAVIRINDQVLPYDKRQPVFEDMTDIDQASAPQ  
RLIIMYKDSSEVRGLAVTLVKDSKMSYLSCKNKIISFEEMDPENIDDIQSDLIFFQ  
KRVPGHNKMEFESSLYEGHFLACQKEDAFKLIKKDKDENGDKSVNFTLTNLHQS"

BASE COUNT 148 a 100 c 130 g 93 t  
ORIGIN

Query Match 90.6%; Score 15.4; DB 56; Length 471;  
Best Local Similarity 76.5%; Pred. No. 4.5e+02;  
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYGARGARATGAYCC 17  
||:||||:||||:|  
Db 244 TTTGAGGAATGATGCC 260

RESULT 11  
MMU66244 572 bp mRNA ROD 18-MAR-1997  
LOCUS Mus musculus interferon-gamma inducing factor mRNA, partial cds.  
DEFINITION U66244  
ACCESSION U66244  
VERSION U66244.1 GI:1561735  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 572)  
AUTHORS Rothe, H., Jenkins, N.A., Copeland, N.G. and Kolb, H.  
TITLE Active stage of autoimmune diabetes is associated with the  
expression of a novel cytokine, IGIF, which is located near Idd2  
JOURNAL J. Clin. Invest. 99 (3), 469-474 (1997)  
MEDLINE 97174346  
REFERENCE 2 (bases 1 to 572)  
AUTHORS Rothe, H., Copeland, N.G. and Kolb, H.  
TITLE Direct Submission  
JOURNAL Submitted (06-AUG-1996) Diabetes Research Institute, Auf'm  
Hennekamp 65, Dueseldorf 40225, Germany

FEATURES  
source  
1..572  
/organism="Mus musculus"  
/strain="non obese diabetic (NOD)"  
/db\_xref="taxon:10090"  
/chromosome="9"  
/tissue\_type="pancreas"



```

CDS
1..>572
/note="IGIF"
/codon_start=1
/product="interferon gamma inducing factor"
/protein_id="AAB49753.1"
/db_xref="GI:1561736"
/translation="MAAMSEDSCVNFKEMFIDNTLYFIPEENGDLSEDNFGRHCTT
AVIRINDQVLEVDKRPVEFEDMDIDQASAEQRLIYMKDSEVRGLAVTLLSKD
SKMSTLSCKNKIISFEEDMPDENIDDIQSDLIFFQKRVPGHNKMEFESSLYEGHFLAC
QKEDDAFKLILKKDENGDKSVISLTNLHQ"
BASE COUNT
196 a 111 c 113 g 152 t
ORIGIN

```

Query Match	90.68;	Score 15.4;	DB 94;	Length 572;
Best Local Similarity	76.58;	Pred. No. 4.6e+02;		
Matches 13; Conservative	4;	Mismatches 0;	Indels 0;	Gaps 0;

OY 1 TTYGARGARATGGAYCC 17  
||:|:|:|:|:|:|:|  
Db 349 TTTGAGGAATGGATCC 365

RESULT	12
MUSIGIFPP	
LOCUS	MUSIGIFPP 866 bp mRNA ROD 10-FEB-1999
DEFINITION	Mouse mRNA for IGF precursor polypeptide, complete cds.
ACCESSION	D49949
VERSION	D49949.1 GI:1064822 1
KEYWORDS	IGF precursor polypeptide.
SOURCE	Mus musculus liver cDNA to mRNA, clone pmuGF37B-5.
ORGANISM	Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 866)

TITLE Direct Submission  
JOURNAL Submitted (29-MAR-1995) to the DDBJ/EMBL/GenBank databases. Haruki

REFERENCE  
2 (bases 1 to 866)  
Okamura, H., Tsutui, H., Komatsu, T., Yutsudo, M., Hakura, A.,  
AUTHORS

**TITLE**  
Cloning of a new cytokine that induces IFN-gamma production by T

JOURNAL Nature 378 (6552), 88-91 (1995)  
MEDLINE 96061009

FEATURES	Location/Qualifiers
source	1. .866

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="liver"
1. .866
mRNA

```

	/note="Interferon-gamma inducing factor (IGIF) mRNA"
CDS	165..743

polyA_site	262 a	187 c	187 g	230 t
BASE COUNT				
ORIGIN				

```

/codon_start=1
/product="IGIF precursor polypeptide"
/protein_id="BAA08705.1"
/db_xref="GI:1064823"
/translation="MAAMSEDCSVNFKEMFIDNTLYFIP EENGDL ESDNFGRLHCTT
AVIRINDOVLFVDKQPVFEDMTIDIGASAPQTRLIIYMKDSEYRGIAVTL SVKD
SKMSFTLSCNKKIISFEEMDPENIDIDSLIFQKRVPGHNKMEFESLLYEGHFLAC
QKEDDAFKLLKKDKDENGKSVMTLTNLHQ$"
866

```

Query Match	90.68;	Score 15.4;	DB 94;	Length 866;
Best Local Similarity	76.58;	Pred. No. 4.7e+02;		
Matches 13; Conservative	4;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	TTYGARGARATGGAYCC	17
	: : : : : :		
Db	513	TTTGAGGAAATGCATCC	529

RESULT	13				
AP000229/c					
LOCUS	AP000229	35143 bp	DNA	PRI	17-MAR-2000
DEFINITION	Homo sapiens genomic DNA, chromosome 21q21.1-q21.2, clone:RL539, LL56-ApF region, complete sequence.				

ACCESSION	AP000229
VERSION	AP000229.2
KEYWORDS	HTG.
SOURCE	Homo sapiens DNA, clone:TI539

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 35143)

**AUTHORS** Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Homma, Y., Fujiyama, A., Yada, T., Totoki, Y. and Sakaki, Y.

**TITLE** Homo sapiens 35,143 genomic DNA of 21q21.1-q21.2

**JOURNAL** Published Only in Database (1999) In press

REFERENCE	2 (bases 1 to 35143)
AUTHORS	Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y. and Sakaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (13-MAY-1999) to the DDBJ/EMBL/GenBank databases.

Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan  
(E-mail: hattori@gsc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-42-778-9923, Fax: 81-42-778-9924)  
On Mar 17, 2000 this sequence version replaced gi:4835598.  
Sequence updated (15-Mar-2000)

The sequencing project is supported by Japan Science Technology Corporation (JST) and The Institute of Physical and Chemical Research (RIKEN).

```
FEATURES      Location/Qualifiers
source        1. .35143
```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="T1539"

```

	/map="21q21.1-q21.2"
BASE COUNT	10208 a 7036 c 7565 g 10334 t
ORIGIN	

Query Match	90.68;	Score 15.4;	DB 90;	Length 35143;
Best Local Similarity	76.58;	Pred. No. 6.5e+02;		
Matches 13; Conservative	4;	Mismatches 0;	Indels 0;	Gaps 0;

```
Qy 1 TTYGARGARATGCAYC 17
    ||:|:|:|:|:|:|:|
Db 25608 TTTGAGGAATGCATCC 25592
```

RESULT 14  
AC080029

LOCUS	AC080029	59955 bp	DNA	HTG	25-OCT-2000
DEFINITION	Homo sapiens chromosome 4 clone CTD-2353K2 map 4, LOW-PASS SEQUENCE SAMPLING.				

ACCESSION	AC080029
VERSION	AC080029.1
KEYWORDS	HTG; HTGS_PHASE0.
SOURCE	human.

ORGANISM	
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
1 (bases 1 to 59955)	
Birren, B., Linton, L., Nusbaum, C. and Lander, E.	

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Homo sapiens chromosome 4, clone CTD-2353K2  
Unpublished  
2 (bases 1 to 59955)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivar, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L10743

Center clone name: 2353\_K\_2

NOTE: This record contains 74 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 645: contig of 645 bp in length  
\* 646 745: gap of 100 bp  
\* 746 1464: contig of 719 bp in length  
\* 1465 1564: gap of 100 bp  
\* 1565 2315: contig of 751 bp in length  
\* 2316 2415: gap of 100 bp  
\* 2416 3059: contig of 644 bp in length  
\* 3060 3159: gap of 100 bp  
\* 3160 3877: contig of 718 bp in length  
\* 3878 3977: gap of 100 bp  
\* 3978 4722: contig of 745 bp in length  
\* 4723 4822: gap of 100 bp  
\* 4823 5579: contig of 757 bp in length  
\* 5580 5679: gap of 100 bp  
\* 5680 6315: contig of 636 bp in length  
\* 6316 6415: gap of 100 bp  
\* 6416 7280: contig of 865 bp in length  
\* 7281 7380: gap of 100 bp  
\* 7381 8134: contig of 754 bp in length  
\* 8135 8234: gap of 100 bp  
\* 8235 8980: contig of 746 bp in length  
\* 8981 9080: gap of 100 bp  
\* 9081 9763: contig of 683 bp in length  
\* 9764 9863: gap of 100 bp

9864 10479: contig of 616 bp in length  
\* 10480 10579: gap of 100 bp  
\* 10580 11203: contig of 624 bp in length  
\* 11204 11303: gap of 100 bp  
\* 11304 12043: contig of 740 bp in length  
\* 12044 12143: gap of 100 bp  
\* 12144 12846: contig of 703 bp in length  
\* 12847 12946: gap of 100 bp  
\* 12947 13678: contig of 732 bp in length  
\* 13679 13778: gap of 100 bp  
\* 13779 14542: contig of 764 bp in length  
\* 14543 14642: gap of 100 bp  
\* 14643 15366: contig of 724 bp in length  
\* 15367 15466: gap of 100 bp  
\* 15467 16160: contig of 694 bp in length  
\* 16161 16260: gap of 100 bp  
\* 16261 17065: contig of 805 bp in length  
\* 17066 17165: gap of 100 bp  
\* 17166 17914: contig of 749 bp in length  
\* 17915 18014: gap of 100 bp  
\* 18015 18715: contig of 701 bp in length  
\* 18716 18815: gap of 100 bp  
\* 18816 19436: contig of 621 bp in length  
\* 19437 19536: gap of 100 bp  
\* 19537 20176: contig of 640 bp in length  
\* 20177 20276: gap of 100 bp  
\* 20277 21013: contig of 737 bp in length  
\* 21014 21113: gap of 100 bp  
\* 21114 21872: contig of 759 bp in length  
\* 21873 21972: gap of 100 bp  
\* 21973 22634: contig of 662 bp in length  
\* 22635 22734: gap of 100 bp  
\* 22735 23340: contig of 606 bp in length  
\* 23341 23440: gap of 100 bp  
\* 23441 24159: contig of 719 bp in length  
\* 24160 24259: gap of 100 bp  
\* 24260 24993: contig of 734 bp in length  
\* 24994 25093: gap of 100 bp  
\* 25094 25806: contig of 713 bp in length  
\* 25807 25906: gap of 100 bp  
\* 25907 26519: contig of 613 bp in length  
\* 26520 26619: gap of 100 bp  
\* 26620 27253: contig of 634 bp in length  
\* 27254 27353: gap of 100 bp  
\* 27354 28058: contig of 705 bp in length  
\* 28059 28158: gap of 100 bp  
\* 28159 28834: contig of 676 bp in length  
\* 28835 28934: gap of 100 bp  
\* 28935 29594: contig of 660 bp in length  
\* 29595 29694: gap of 100 bp  
\* 29695 30458: contig of 764 bp in length  
\* 30459 30558: gap of 100 bp  
\* 30559 31323: contig of 765 bp in length  
\* 31324 31423: gap of 100 bp  
\* 31424 32170: contig of 747 bp in length  
\* 32171 32270: gap of 100 bp  
\* 32271 33037: contig of 767 bp in length  
\* 33038 33137: gap of 100 bp  
\* 33138 33845: contig of 708 bp in length  
\* 33846 33945: gap of 100 bp  
\* 33946 34637: contig of 692 bp in length  
\* 34638 34737: gap of 100 bp  
\* 34738 35427: contig of 690 bp in length  
\* 35428 35527: gap of 100 bp  
\* 35528 36173: contig of 646 bp in length  
\* 36174 36273: gap of 100 bp  
\* 36274 36921: contig of 648 bp in length  
\* 36922 37021: gap of 100 bp  
\* 37022 37736: contig of 715 bp in length  
\* 37737 37836: gap of 100 bp  
\* 37837 38624: contig of 788 bp in length  
\* 38625 38724: gap of 100 bp  
\* 38725 39480: contig of 756 bp in length

```

* 39481 39580: gap of 100 bp
* 39581 40327: contig of 747 bp in length
* 40328 40427: gap of 100 bp
* 40428 41188: contig of 761 bp in length
* 41189 41288: gap of 100 bp
* 41289 42060: contig of 772 bp in length
* 42061 42160: gap of 100 bp
* 42161 42894: contig of 734 bp in length
* 42895 42994: gap of 100 bp
* 42995 43747: contig of 753 bp in length
* 43748 43847: gap of 100 bp
* 43848 44462: contig of 615 bp in length
* 44463 44562: gap of 100 bp
* 44563 45197: contig of 635 bp in length
* 45198 45297: gap of 100 bp
* 45298 46034: contig of 737 bp in length
* 46035 46134: gap of 100 bp
* 46135 46891: contig of 757 bp in length
* 46892 46991: gap of 100 bp
* 46992 47697: contig of 706 bp in length
* 47698 47797: gap of 100 bp
* 47798 48552: contig of 755 bp in length
* 48553 48652: gap of 100 bp
* 48653 49293: contig of 641 bp in length
* 49294 49393: gap of 100 bp
* 49394 50099: contig of 706 bp in length
* 50100 50199: gap of 100 bp
* 50200 50957: contig of 758 bp in length
* 50958 51057: gap of 100 bp
* 51058 51840: contig of 783 bp in length
* 51841 51940: gap of 100 bp
* 51941 52645: contig of 705 bp in length
* 52646 52745: gap of 100 bp
* 52746 53371: contig of 626 bp in length
* 53372 53471: gap of 100 bp
* 53472 54088: contig of 617 bp in length
* 54089 54188: gap of 100 bp
* 54189 54963: contig of 775 bp in length
* 54964 55063: gap of 100 bp
* 55064 55797: contig of 734 bp in length
* 55798 55897: gap of 100 bp
* 55898 56622: contig of 725 bp in length

```

```

Query Match      90.6%; Score 15.4; DB 77; Length 59955;
Best Local Similarity 76.5%; Pred. No. 6.8e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 TTYGARGARATGAYCC 17
    11:11:11:11:11:11
Db 55674 TTTGAAGAATGACCC 55690

```

```

RESULT 15
AC078950/c 95155 bp DNA PRI 01-JAN-2001
LOCUS Homo sapiens 12 BAC RP11-407P10 (Roswell Park Cancer Institute
DEFINITION Human BAC Library) complete sequence.
ACCESSION AC078950
VERSION AC078950.19 GI:12000438
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM human.

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 95155)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blumage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
COMMENT

```

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,  
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,  
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,  
Gunnarone,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,  
Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,  
Hodgson,A., Hogues,M., Hollaway,C., Hollins,B., Homsif.,  
Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapa,P., Marondel,I., Martin,R.,  
Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P.,  
Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,  
Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,  
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S.,  
Oguh,M., Okwunu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,  
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,  
Shoostari,N., Slisson,I., Sodergren,E., Sonaiki,T., Sparks,A.,  
Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,  
Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,  
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D.,  
Washington,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,  
Washington,C., Watlington,S., Williams,G., Williamson,A.,  
Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,  
Zorrilla,S., Kucherlapati,R. and Gibbs,R.

Unpublished  
2 (bases 1 to 95155)  
Worley,K.C.  
Direct Submission  
Submitted (13-AUG-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jan 1, 2001 this sequence version replaced gi:12000225.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTSs, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2001, 10:13:29 ; Search time 1105.85 Seconds  
(without alignments)  
226.716 Million cell updates/sec

Title: US-09-373-230-5  
Perfect score: 17  
Sequence: 1 TTYGARGARATGAYCC 17

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1283235 segs, 7373929652 residues  
Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_com:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgo\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_com:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_vi:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_vil:\*  
59: gb\_vil2:\*  
60: gb\_vil3:\*  
61: gb\_vil4:\*  
62: gb\_vil5:\*  
63: gb\_vil6:\*  
64: gb\_vil7:\*  
65: gb\_vil8:\*  
66: gb\_vil9:\*  
67: gb\_vil10:\*  
68: gb\_vil11:\*  
69: gb\_vil12:\*  
70: gb\_vil13:\*  
71: gb\_vil14:\*  
72: gb\_vil15:\*  
73: gb\_vil16:\*  
74: gb\_vil17:\*  
75: gb\_vil18:\*  
76: gb\_vil19:\*  
77: gb\_vil20:\*  
78: gb\_vil21:\*  
79: gb\_vil22:\*  
80: gb\_vil23:\*  
81: gb\_vil24:\*  
82: gb\_vil25:\*  
83: gb\_vil26:\*  
84: gb\_vil27:\*  
85: gb\_vil28:\*  
86: gb\_vil29:\*  
87: gb\_vil30:\*  
88: gb\_vil31:\*  
89: gb\_vil32:\*  
90: gb\_vil33:\*  
91: gb\_vil34:\*  
92: gb\_vil35:\*  
93: gb\_vil36:\*  
94: gb\_vil37:\*  
95: gb\_vil38:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.4	90.6	17	9 AR072047	AR072047 Sequence
2	15.4	90.6	471	9 AR072044	AR072044 Sequence
3	15.4	90.6	471	10 E13264	E13264 Mouse cDNA
4	15.4	90.6	471	10 E14257	E14257 cDNA encodi
5	15.4	90.6	471	10 E14760	E14760 cDNA encodi
6	15.4	90.6	471	10 E17139	E17139 Murine mRNa
7	15.4	90.6	471	45 E10609	E10609 Mouse cDNA
8	15.4	90.6	471	45 E11744	E11744 cDNA encodi
9	15.4	90.6	471	45 E12010	E12010 cDNA encodi
10	15.4	90.6	471	56 ASIGIF	Y09278 Artificial
11	15.4	90.6	572	94 MM066244	U66244 Mus musculu

12	15.4	90.6	866	94	MUSIGIFPP	D49949 Mouse mRNA
13	15.4	90.6	35143	90	AP000229	AP000229 Homo sapi
14	15.4	90.6	59955	77	AC080029	AC080029 Homo sapi
15	15.4	90.6	95155	88	AC078950	AC078950 Homo sapi
16	15.4	90.6	100000	90	AP000144	AP000144 Homo sapi
17	15.4	90.6	100634	90	AP001594	AP001594 Homo sapi
18	15.4	90.6	123631	91	HS22F01	AL109967 Homo sapi
19	15.4	90.6	202103	87	AC010176	AC010176 Homo sapi
20	15.4	90.6	340000	90	AP001695	AP001695 Homo sapi
21	14.4	84.7	355	54	G01707	G01707 chicken STS
22	14.4	84.7	710	94	MMBEGFL4	U39192 Mus musculu
23	14.4	84.7	963	53	CNS01GZU	AL143803 Anopheles
24	14.4	84.7	997	54	CNS07CNO	AL439404 T7 end of
25	14.4	84.7	1084	53	CNS07OT9	AL424051 clone AZ0
26	14.4	84.7	1188	5	AF318610	AF318610 Caenorhab
27	14.4	84.7	1718	12	AF046922	AF046922 Colletotr
28	14.4	84.7	1971	4	AB035057	AB035057 Drosophill
29	14.4	84.7	2094	4	AB035066	AB035066 Drosophill
30	14.4	84.7	2178	94	MUSHBEGF04	L36027 Mus musculu
31	14.4	84.7	2237	4	AB035069	AB035069 Drosophill
32	14.4	84.7	2268	4	AB035062	AB035062 Drosophill
33	14.4	84.7	2282	4	AB035061	AB035061 Drosophill
34	14.4	84.7	2305	4	AB035065	AB035065 Drosophill
35	14.4	84.7	2919	12	AF024633	AF024633 Magnaport
36	14.4	84.7	3172	12	AF015753	AF015753 Magnaport
37	14.4	84.7	3522	4	AB035058	AB035058 Drosophill
38	14.4	84.7	4438	8	AF085184	AF085184 Gallus ga
39	14.4	84.7	13952	13	RICPFA	D32136 Rice endoge
40	14.4	84.7	20128	14	AF155773	AF155773 Gibberell
41	14.4	84.7	21077	5	CEK10C8	Z74474 Caenorhabdi
42	14.4	84.7	21644	92	HS22A1KLT	Z81330 Human DNA s
43	14.4	84.7	22092	4	AC006637	AC006637 Caenorhab
44	14.4	84.7	24361	5	CELRO2C2	AF039037 Caenorhab
45	14.4	84.7	27748	5	CEIH34C03	AF100662 Caenorhab

## ALIGNMENTS

RESULT 1

LOCUS AR072047 17 bp DNA PAT 18-FEB-2000

DEFINITION Sequence 5 from patent US 5912324.

ACCESSION AR072047

VERSION AR072047.1 GI:7222935

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

17; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

cdna encoding polypeptide which induces interferon-gamma production  
HIGIF.

```
CC strandedness: Double;
CC topology: Linear;
```

```

FH      Key      Location/Qualifiers
FT      source      1..471
FT      /organism='Mus sp.'
FT      /tissue_type='liver'
FT      mat_peptide 1..471
FT      /product='IL-18'
FEATURES
source      1..471
            /organism='Mus sp.'
            /db_xref='taxon:10095'
BASE COUNT  162 a 91 c 92 g 126 t
ORIGIN
Query Match      90.6%; Score 15.4; DB 10; Length 471;
Best Local Similarity 76.5%; Pred. No. 4.5e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTYGARGARATGAYCC 17
        ||:||||:||||:|
DB      244 TTGAGGAATGATCC 260

RESULT  7
E10609  standard; RNA; ROD; 471 BP.
XX      E10609;
XX      E10609.1
SV      E10609.1
XX      08-OCT-1997 (Rel. 52, Created)
DT      02-SEP-2000 (Rel. 65, last updated, Version 2)
XX      Mouse cDNA encoding a protein involved in interferon-gamma production.
DE      JP 1996027189-A/1.
XX      Mus sp.
XX      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX      [1]
RN      1-471
RP      Okamura H., Tanimoto T., Torigoe K., Kurimoto M.;
RA      "PROTEIN INDUCING PRODUCTION OF INTERFERON-GAMMA";
RT      Patent number JP1996027189-A/1, 30-JAN-1996.
RL      HAYASHIBARA BIOCHEM LAB INC.
XX      OS
XX      Mus sp. (mouse)
CC      PN JP 1996027189-A/1
CC      PD 30-JAN-1996
CC      PF 14-JUL-1994 JP 1994184162
CC      PI OKAMURA HARUKI, TANIMOTO TADAO, TORIGOE KAKUTI,
CC      KURIMOTO MASASHI
CC      PC C07K14/52,A61K38/00,A61K38/00,C12N1/21,C12N15/09,
CC      C12P21/02//C07K14/57;
CC      PC strandedness: Double;
CC      CC topology: Linear;
CC      CC Key
CC      FH Location/Qualifiers
CC      FT source
CC      FT 1..471
CC      FT /organism='Mus sp.'
CC      FT /tissue_type='liver'
CC      FT mat_peptide
CC      FT 1..471
CC      FT /product='a protein involved in
CC      FT interferon-gamma
CC      FT production"
XX      Key
XX      Location/Qualifiers
FH      Key
FH      Location/Qualifiers
FT      source      1..471

```

```

FT      /db_xref='taxon:10095'
FT      /organism='Mus sp.'
XX      SQ      Sequence 471 BP; 162 A; 91 C; 92 G; 125 T; 1 other;
Query Match      90.6%; Score 15.4; DB 45; Length 471;
Best Local Similarity 76.5%; Pred. No. 4.5e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTYGARGARATGAYCC 17
        ||:||||:||||:|
DB      244 TTGAGGAATGATCC 260

RESULT  8
E11744  standard; RNA; ROD; 471 BP.
XX      E11744;
XX      E11744.1
SV      E11744.1
XX      08-OCT-1997 (Rel. 52, Created)
DT      02-SEP-2000 (Rel. 65, last updated, Version 2)
XX      cDNA encoding polypeptide which induce mouse interferon-gamma product.
DE      JP 1996193098-A/1.
XX      Mus musculus (house mouse)
XX      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX      [1]
RN      1-471
RP      Ushio S., Torigoe K., Tanimoto T., Okamura H., Kurimoto M.;
RA      "POLYPEPTIDE FOR INDUCING PRODUCTION OF INTERFERON-GAMMA";
RT      Patent number JP1996193098-A/1, 30-JUL-1996.
RL      HAYASHIBARA BIOCHEM LAB INC.
XX      OS
XX      Mus musculus (mouse)
CC      PN JP 1996193098-A/1
CC      PD 30-JUL-1996
CC      PF 18-SEP-1995 JP 1995262062
CC      PR 15-NOV-1994 JP 94P 304203
CC      PI USHIO SHINPEI, TORIGOE KAKUTI, TANIMOTO TADAO, OKAMURA HARUKI,
CC      KURIMOTO MASASHI
CC      PC C07K14/52,C07H21/04,C12N1/21,C12N15/09,C12P21/00//A61K38/00,
CC      C07K14/06,
CC      PC C07K7/08, (C12N1/21,C12R1:19), (C12P21/00,C12R1:19);
CC      CC strandedness: Double;
CC      CC topology: Linear;
CC      CC hypothetical: No;
CC      CC anti-sense: No;
CC      CC Key
CC      FH Location/Qualifiers
CC      FT source
CC      FT 1..471
CC      FT /db_xref='taxon:10090'
CC      FT /organism='Mus musculus'
XX      SQ      Sequence 471 BP; 162 A; 91 C; 92 G; 125 T; 1 other;
Query Match      90.6%; Score 15.4; DB 45; Length 471;
Best Local Similarity 76.5%; Pred. No. 4.5e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```



OY 1 TTYGARGARATGAYCC 17  
11:11:11:11:11:11  
Db 244 TTTGAGGAATGATCC 260

RESULT 9  
E12010  
ID E12010 standard; RNA; ROD: 471 BP.  
XX  
AC E12010;  
XX E12010.1  
SV  
XX 07-OCT-1997 (Rel. 52, Created)  
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)  
XX  
DE cDNA encoding mouse polypeptide which introduce interferon-gamma product in  
DE immunocompetent cell.  
XX  
KW JP 1996231598-A/2.  
XX  
OS Mus musculus (house mouse)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
XX  
RN [1]  
RP 1-471  
RA Kunikata T., Taniguchi M., Kono K., Kurimoto M.;  
RT "MONOCLONAL ANTIBODY";  
RL Patent number JP1996231598-A/2, 10-SEP-1996.  
RL HAYASHIBARA BIOCHEM LAB INC.  
XX  
CC OS Mus musculus (mouse)  
CC PN JP 1996231598-A/2  
CC PD 10-SEP-1996  
CC PF 23-FEB-1995 JP 1995058240  
CC PI KUNIKATA TOSHIO, TANIGUCHI MUTSUOKO, KONO KEIZO,  
CC PI KURIMOTO MASASHI  
CC PC C07K16/24,C07K1/16,C07K1/18,C07K1/22,C07K1/26,C07K1/30,  
CC PC C07K1/34,C12N5/10,  
CC PC C12N15/02,C12P21/08,G01N33/53,G01N33/577//A61K38/21,A61K39/395,  
CC PC C12P21/08,  
CC PC C12R1/91;  
CC CC strandedness: Double;  
CC CC topology: Linear;  
CC CC hypothetical: No;  
CC CC anti-sense: No;  
CC FH key Location/Qualifiers  
CC FH source 1..471  
CC FT /organism="Mus musculus"  
CC FT /tissue\_type="liver"  
XX  
FH key Location/Qualifiers  
FH FT source 1..471  
FH FT /db\_xref="taxon:10090"  
FH FT /organism="Mus musculus"  
FT  
XX  
SQ Sequence 471 BP; 162 A; 91 C; 92 G; 125 T; 1 other;

Query Match 90.6%; Score 15.4; DB 45; Length 471;  
Best Local Similarity 76.5%; Pred. No. 4.5e+02;  
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAYCC 17  
11:11:11:11:11:11  
Db 244 TTTGAGGAATGATCC 260

RESULT 10

ASIGIF  
LOCUS 471 bp DNA SYN 08-NOV-1996  
DEFINITION Artificial sequence DNA for cytokine Interferon-gamma inducing factor (IGIF).  
ACCESSION Y09278  
VERSION Y09278.1 GI:1666283  
KEYWORDS cytokine; interferon-gamma inducing factor (IGIF).  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 471)  
AUTHORS Schluesener, H.J.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 471)  
AUTHORS Schluesener, H.J.  
TITLE Direct Submission  
JOURNAL Submitted (21-OCT-1996) H.J. Schluesener, Inst.f. Hirnforschung, Calwer Str. 3, D- 77076 Tuebingen, FRG

FEATURES  
source  
1..471  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
CDS  
1..>471  
/codon\_start=1  
/transl\_table=11  
/product="mouse cytokine interferon-gamma inducing factor"  
/protein\_id="CAA70473.1"  
/db\_xref="GI:1666284"  
/translation="NFGRLHCTTAVIRINDVLFVDRKRPVEEDMTDIDQASAPQ  
RLIYMKDSEVRGLAVTLVSKDSKSTLSCKNKIISFEEDPENNIDDIQSDLIFFQ  
KRVPGHNKMEFESSLYEGHFLACQEDDAFKILKKKDENDKSVMTLTNLHQS"

BASE COUNT 148 a 100 c 130 g 93 t  
ORIGIN

Query Match 90.6%; Score 15.4; DB 56; Length 471;  
Best Local Similarity 76.5%; Pred. No. 4.5e+02;  
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAYCC 17  
11:11:11:11:11:11  
Db 244 TTTGAGGAATGATCC 260

RESULT 11  
MMU66244  
LOCUS MMU66244 572 bp mRNA ROD 18-MAR-1997  
DEFINITION Mus musculus interferon-gamma inducing factor mRNA, partial cds.  
ACCESSION U66244  
VERSION U66244.1 GI:1561735  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE 1 (bases 1 to 572)  
AUTHORS Rothe, H., Jenkins, N.A., Copeland, N.G. and Kolb, H.  
TITLE Active stage of autoimmune diabetes is associated with the expression of a novel cytokine, IGIF, which is located near Idd2  
JOURNAL J. Clin. Invest. 99 (3), 469-474 (1997)  
MEDLINE 97174346  
REFERENCE 2 (bases 1 to 572)  
AUTHORS Rothe, H., Copeland, N.G. and Kolb, H.  
TITLE Direct Submission  
JOURNAL Submitted (06-AUG-1996) Diabetes Research Institute, Auf'm Hennekamp 65, Duesseldorf 40225, Germany

FEATURES  
source  
1..572  
/organism="Mus musculus"  
/strain="non obese diabetic (NOD)"  
/db\_xref="taxon:10090"  
/chromosome="9"  
/tissue\_type="pancreas"

```
CDS
1..>572
/note="IGIF"
/codon_start=1
/product="interferon gamma inducing factor"
/protein_id="AAB49753.1"
/db_xref="GI:1561736"
/translation="MAAMSEDSVCNFKEMFIDNTLYFIPEENGDLSDNFGRLHCTT
AVIRNINDOVLFDKROPVDEMDIDOSASEPOTRLIIMYKDSVGRGLAVTLISKD
SKMSTLSCKNKRIISFEEMDPENIDDIOSDLIFFQKRVPGHNKMEFESSLIEGHFLAC
OKEDDAFKLILKKDKDENGKSVISLTNLHQ"
BASE COUNT      196 a      111 c      113 g      152 t
ORIGIN

Query Match      90.6%; Score 15.4; DB 94; Length 572;
Best Local Similarity 76.5%; Pred. No. 4.6e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTYGARGARATGAYCC 17
      ||:||||:||||:|
Db      349 TTTGAGGAATGATCC 365

RESULT 12
MUSIGIFP      866 bp      mRNA      ROD      10-FEB-1999
LOCUS      Mouse mRNA for IGIF precursor polypeptide, complete cds.
ACCESSION      D49949
VERSION      D49949.1 GI:1064822
KEYWORDS      IGIF precursor polypeptide.
SOURCE      Mus musculus liver cDNA to mRNA, clone pMUGF37B-5.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 866)
AUTHORS      Okamura,H.
TITLE      Direct Submission
JOURNAL      Submitted (29-MAR-1995) to the DDBJ/EMBL/GenBank databases. Haruki
      Okamura, Hyogo College of Medicine, Department of Bacteriology;
      Mukogawa I-1, Nishinomiya, Hyogo 663, Japan (Tel:0798-45-6111)
2 (bases 1 to 866)
REFERENCE      Okamura,H., Tsutui,H., Komatsu,T., Yutsudo,M., Hakura,A.,
      Tanimoto,T., Torigoe,K., Okura,T., Nakada,Y., Hattori,K., Akita,K.,
      Namba,M., Tanabe,F., Konishi,K., Fukuda,S. and Kurimoto,M.
TITLE      Cloning of a new cytokine that induces IFN-gamma production by T
      cells
JOURNAL      Nature 378 (6552), 88-91 (1995)
MEDLINE      96061009
FEATURES
      source      Location/Qualifiers
      1..866
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /tissue_type="liver"
      1..866
      /note="Interferon-gamma inducing factor (IGIF) mRNA"
      165..743
      /codon_start=1
      /product="IGIF precursor polypeptide"
      /protein_id="BAA08705.1"
      /db_xref="GI:1064823"
      /translation="MAAMSEDSVCNFKEMFIDNTLYFIPEENGDLSDNFGRLHCTT
      AVIRNINDOVLFDKROPVDEMDIDOSASEPOTRLIIMYKDSVGRGLAVTLISKD
      SKMSTLSCKNKRIISFEEMDPENIDDIOSDLIFFQKRVPGHNKMEFESSLIEGHFLAC
      OKEDDAFKLILKKDKDENGKSVISLTNLHQ"
      866
polyA_site      262 a      187 c      187 g      230 t
BASE COUNT
ORIGIN

Query Match      90.6%; Score 15.4; DB 94; Length 866;
Best Local Similarity 76.5%; Pred. No. 4.7e+02;
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTYGARGARATGAYCC 17
      ||:||||:||||:|
Db      25608 TTTGAGGAATGATCC 25592

RESULT 14
AC080029      5995 bp      DNA      HTG      25-OCT-2000
LOCUS      Homo sapiens chromosome 4 clone CTD-2353K2 map 4, LOW-PASS SEQUENCE
DEFINITION      SAMPLING.
ACCESSION      AC080029
VERSION      AC080029.1 GI:10280784
KEYWORDS      HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 5995)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
```

TITLE  
JOURNAL  
REFERENCE  
AUTHORSHomo sapiens chromosome 4, clone CTD-2353K2  
Unpublished  
2 (bases 1 to 59955)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArrelano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Harford, A., Horton, L., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Larocque, K., Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Lieu, C., Liu, G., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M., Oliver, J., Peterson, K., Pierre, N., Pisanic, C., Pollara, V., Raymond, C., Rieback, M., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L10743

Center clone name: 2353\_K\_2

NOTE: This record contains 74 individual  
sequencing reads that have not been assembled into  
contigs. Runs of N are used to separate the reads  
and the order in which they appear is completely  
arbitrary. Low-pass sequence sampling is useful for  
identifying clones that may be gene-rich and allows  
overlap relationships among clones to be deduced.  
However, it should not be assumed that this clone  
will be sequenced to completion. In the event that  
the record is updated, the accession number will  
be preserved.

1 645: contig of 645 bp in length  
646 745: gap of 100 bp  
746 1464: contig of 719 bp in length  
1465 1564: gap of 100 bp  
1565 2315: contig of 751 bp in length  
2316 2415: gap of 100 bp  
2416 3059: contig of 644 bp in length  
3060 3159: gap of 100 bp  
3160 3877: contig of 718 bp in length  
3878 3977: gap of 100 bp  
3978 4722: contig of 745 bp in length  
4723 4822: gap of 100 bp  
4823 5579: contig of 757 bp in length  
5580 5679: gap of 100 bp  
5680 6315: contig of 636 bp in length  
6316 6415: gap of 100 bp  
6416 7280: contig of 865 bp in length  
7281 7380: gap of 100 bp  
7381 8134: contig of 754 bp in length  
8135 8234: gap of 100 bp  
8235 8980: contig of 746 bp in length  
8981 9080: gap of 100 bp  
9081 9763: contig of 683 bp in length  
9764 9863: gap of 100 bp

9864 10479: contig of 616 bp in length  
10480 10579: gap of 100 bp  
10580 11203: contig of 624 bp in length  
11204 11303: gap of 100 bp  
11304 12043: contig of 740 bp in length  
12044 12143: gap of 100 bp  
12144 12846: contig of 703 bp in length  
12847 12946: gap of 100 bp  
12947 13678: contig of 732 bp in length  
13679 13778: gap of 100 bp  
13779 14542: contig of 764 bp in length  
14543 14642: gap of 100 bp  
14643 15366: contig of 724 bp in length  
15367 15466: gap of 100 bp  
15467 16160: contig of 694 bp in length  
16161 16260: gap of 100 bp  
16261 17065: contig of 805 bp in length  
17066 17165: gap of 100 bp  
17166 17914: contig of 749 bp in length  
17915 18014: gap of 100 bp  
18015 18715: contig of 701 bp in length  
18716 18815: gap of 100 bp  
18816 19436: contig of 621 bp in length  
19437 19536: gap of 100 bp  
19537 20176: contig of 640 bp in length  
20177 20276: gap of 100 bp  
20277 21013: contig of 737 bp in length  
21014 21113: gap of 100 bp  
21114 21872: contig of 759 bp in length  
21873 21972: gap of 100 bp  
21973 22634: contig of 662 bp in length  
22635 22734: gap of 100 bp  
22735 23340: contig of 606 bp in length  
23341 23440: gap of 100 bp  
23441 24159: contig of 719 bp in length  
24160 24259: gap of 100 bp  
24260 24993: contig of 734 bp in length  
24994 25093: gap of 100 bp  
25094 25806: contig of 713 bp in length  
25807 25906: gap of 100 bp  
25907 26519: contig of 613 bp in length  
26520 26619: gap of 100 bp  
26620 27253: contig of 634 bp in length  
27254 27353: gap of 100 bp  
27354 28058: contig of 705 bp in length  
28059 28158: gap of 100 bp  
28159 28834: contig of 676 bp in length  
28835 28934: gap of 100 bp  
28935 29594: contig of 660 bp in length  
29595 29694: gap of 100 bp  
29695 30458: contig of 764 bp in length  
30459 30558: gap of 100 bp  
30559 31323: contig of 765 bp in length  
31324 31423: gap of 100 bp  
31424 32170: contig of 747 bp in length  
32171 32270: gap of 100 bp  
32271 33037: contig of 767 bp in length  
33038 33137: gap of 100 bp  
33138 33845: contig of 708 bp in length  
33846 33945: gap of 100 bp  
33946 34637: contig of 692 bp in length  
34638 34737: gap of 100 bp  
34738 35427: contig of 690 bp in length  
35428 35527: gap of 100 bp  
35528 36173: contig of 646 bp in length  
36174 36273: gap of 100 bp  
36274 36921: contig of 648 bp in length  
36922 37021: gap of 100 bp  
37022 37736: contig of 715 bp in length  
37737 37836: gap of 100 bp  
37837 38624: contig of 788 bp in length  
38625 38724: gap of 100 bp  
38725 39480: contig of 756 bp in length





QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

## Summary Statistics

Contig length:	173652
Phrap values in estimate:	170386
Average error rate (BCM-Phrap estimate):	0.000194024
Fraction of Phrap values less than 40 :	0.0280363
Number of consensus changing edits:	1401
Number of N's in consensus :	0

Position	Consensus	Changing	edits	Edited+Context
1	.....	.....	.....	.....
2	.....	.....	.....	.....
3	.....	.....	.....	.....
4	.....	.....	.....	.....
5	.....	.....	.....	.....
6	.....	.....	.....	.....
7	.....	.....	.....	.....
8	.....	.....	.....	.....
9	.....	.....	.....	.....
10	.....	.....	.....	.....
11	.....	.....	.....	.....
12	.....	.....	.....	.....
13	.....	.....	.....	.....
14	.....	.....	.....	.....
15	.....	.....	.....	.....
16	.....	.....	.....	.....
17	.....	.....	.....	.....
18	.....	.....	.....	.....
19	.....	.....	.....	.....
20	.....	.....	.....	.....
21	.....	.....	.....	.....
22	.....	.....	.....	.....
23	.....	.....	.....	.....
24	.....	.....	.....	.....
25	.....	.....	.....	.....
26	.....	.....	.....	.....
27	.....	.....	.....	.....
28	.....	.....	.....	.....
29	.....	.....	.....	.....
30	.....	.....	.....	.....
31	.....	.....	.....	.....
32	.....	.....	.....	.....
33	.....	.....	.....	.....
34	.....	.....	.....	.....
35	.....	.....	.....	.....
36	.....	.....	.....	.....
37	.....	.....	.....	.....
38	.....	.....	.....	.....
39	.....	.....	.....	.....
40	.....	.....	.....	.....
41	.....	.....	.....	.....
42	.....	.....	.....	.....
43	.....	.....	.....	.....
44	.....	.....	.....	.....
45	.....	.....	.....	.....
46	.....	.....	.....	.....
47	.....	.....	.....	.....
48	.....	.....	.....	.....

49	nnnnnnnnnn(n)nnnnnnnnnn	gttatggtc(a)atgtctgtaag
50	nnnnnnnnnn(n)nnnnnnnnnn	tttatggtta(a)tgctgtaaga
51	nnnnnnnnnn(n)nnnnnnnnnn	ttatggttaa(t)gtctgtaaga
52	nnnnnnnnnn(n)nnnnnnnnnn	tatgtgttaat(g)tgtaagaac
53	nnnnnnnnnn(n)nnnnnnnnnn	atggttcaatg(t)tgtaagaaca
54	nnnnnnnnnn(n)nnnnnnnnnn	tggtttaatgt(t)gtaagaacat
55	nnnnnnnnnn(n)nnnnnnnnnn	ggtttaatgt(g)taagaacata
56	nnnnnnnnnn(n)nnnnnnnnnn	gttaatgttg(t)aagaacatat
57	nnnnnnnnnn(n)nnnnnnnnnn	ltaatgttgt(a)agaacatatg
58	nnnnnnnnnn(n)nnnnnnnnnn	taatgttgtta(a)gaacatatga
59	nnnnnnnnnn(n)nnnnnnnnnn	aatgttgttaa(g)aacatatgaa
60	nnnnnnnnnn(n)nnnnnnnnnn	atgttgttaag(a)acatatgaa
61	nnnnnnnnnn(n)nnnnnnnnnn	tgittgtaaga(a)acatgaaag
62	nnnnnnnnnn(n)nnnnnnnnnn	gttgttaagaa(c)atatgaaagc
63	nnnnnnnnnn(n)nnnnnnnnnn	ttgtaagaac(a)tatgaaagca
64	nnnnnnnnnn(n)nnnnnnnnnn	tgtaagaaca(t)atgaaagcag
65	nnnnnnnnnn(n)nnnnnnnnnn	gtaagaacat(a)tgaaagcag
66	nnnnnnnnnn(n)nnnnnnnnnn	taagaacata(t)gaaagcagga
67	nnnnnnnnnn(n)nnnnnnnnnn	aagaacatat(g)aaagcagga
68	nnnnnnnnnn(n)nnnnnnnnnn	agaacatatg(a)aagcaggaac
69	nnnnnnnnnn(n)nnnnnnnnnn	gaacatatga(a)agcaggaaca
70	nnnnnnnnnn(n)nnnnnnnnnn	aacatatga(a)ggaaggaacag
71	nnnnnnnnnn(n)nnnnnnnnnn	acatatgaaa(g)caggaaacaga
72	nnnnnnnnnn(n)nnnnnnnnnn	catatgaaag(c)aggaacagaa
73	nnnnnnnnnn(n)nnnnnnnnnn	atatgaaagc(a)ggaacagaat
74	nnnnnnnnnn(n)nnnnnnnnnn	tatgaaagca(g)gaacagaaata
75	nnnnnnnnnn(n)nnnnnnnnnn	atgaaagcag(g)aacagaataa
76	nnnnnnnnnn(n)nnnnnnnnnn	tgaaagcag(g)aacagaataaa
77	nnnnnnnnnn(n)nnnnnnnnnn	gaaagccagga(a)cagaaataaa
78	nnnnnnnnnn(n)nnnnnnnnnn	aaagccagga(c)agaataaaat
79	nnnnnnnnnn(n)nnnnnnnnnn	aagcaggaac(a)gaataaaata
80	nnnnnnnnnn(n)nnnnnnnnnn	agcaggaaca(g)ataaaataa
81	nnnnnnnnnn(n)nnnnnnnnnn	gcaggaacag(a)ataaaataat
82	nnnnnnnnnn(n)nnnnnnnnnn	caggaacaga(a)taaaataatg
83	nnnnnnnnnn(n)nnnnnnnnnn	aggaaccagaa(t)aaaaataatgt

Query Match	90.6%;	Score 15.4;	DB 88;	Length 95155;
Best Local Similarity	76.5%;	Pred. No. 7.1e+02;		
Matches 13;	Conservative	4;	Mismatches 0;	Indels 0;
			Gaps	0

```
QY      1 TTYGARGARATGGAYCC 17
        ||:||:||:||:||:||
Db      69860 TTTGAAGAGATGGATCC 69844
```

RESULT 16  
AP000144/c

LOCUS	AP000144	100000 bp	DNA	PRI	08-JAN-2000
DEFINITION	Homo sapiens genomic DNA, chromosome 21q21.2, LB56-APP region				
DESCRIPTION	clone B2291C14-R44F3, segment 9/10, complete sequence.				

VERSION	AP000144.1	GI:4827110
KEYWORDS	HTG.	
SOURCE	Homo sapiens DNA.	

REFERENCE  
1 (bases 1 to 100000)  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE Fujiyama, A., Yada, T., Totoki, Y. and Sakaki, Y.  
Homo sapiens 911,949bp genomic DNA of 21q21.2 (REGION: L156-APP)  
CLONE RANGE: B2291C14-R44F3)

REFERENCE  
AUTHORS  
2 (bases 1 to 100000)  
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,  
Fujiyama, A., Yada, T., Totoki, Y. and Sakaki, Y.

**JOURNAL**  
Submitted (10-MAY-1999) to the DDBJ/EMBL/Genbank databases.  
Masahira Hattori, The Institute of Physical and Chemical Research  
(RIKEN), Genomic Sciences Center (GSC): Kitasato Univ., 1-15-1  
Kitasato, Sagamihara, Kanagawa 228-8555, Japan  
(E-mail: hattori@gsc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>,  
Tel: 81-42-778-9923, Fax: 81-42-778-9924)







```

VERSION      AC010176.12   GI:10190751
KEYWORDS     HTG.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 202103)
AUTHORS      Mamalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chiu,D., Chowdhury,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Eamhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
Gunnaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsli,F.,
Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
Karlssohn,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
Martindale,A., Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P.,
Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
Shooshtari,N., Slissov,I., Sodergren,E., Sonaike,T., Sparks,A.,
Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Watlington,S., Williams,G., Williamson,A.,
Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
Zorrilla,S., Kucheriapati,R. and Gibbs,R.
TITLE        unpublished
JOURNAL      Direct Submission
AUTHORS      2 (bases 1 to 202103)
REFERENCE    Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (15-SEP-1999) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3 (bases 1 to 202103)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (19-SEP-2000) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Sep 19, 2000 this sequence version replaced g1:9887569.
              INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
              gc-help@bcm.tmc.edu

```

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT

```

----- Summary Statistics -----
Contig length:      202103
Phrap values in estimate: 201669
Average error rate (BCM-Phrap estimate): 9.77954e-05
Fraction of Phrap values less than 40 : 0.0287005
Number of consensus changing edits: 29
Number of N's in consensus : 0

```

Position	Original+Context	Edited+Context
11794	taatctcttt(n)catgcatlcc	taatctcttt(t)catgcatlcc
13909	atggaacaa(n)cttcccatgt	atggaacaa(a)cttcccatgt
39837	agagggcaat(n)latagcacta	agagggcaat(t)latagcacta
39928	aagagcaaac(n)catlcaaaag	aagagcaaac(a)catlcaaaag
45888	cacctgtat(n)taacaacacca	cacctgtat(c)taacaacacca
54727	aagaagaat(n)caaaaatctg	aagaagaat(a)caaaaatctg
54740	aaaatctgtg(n)ccttccctgg	aaaatctgt(g)ccttccctgg
73838	atctcccaaa(n)ctatccctcc	atctcccaaa(g)ctatccctcc
73861	ccctcccccct(n)cccaccacag	ccctcccccct(c)cccaccacag
89489	tacattgat(n)tglatcmtan	tacattgat(t)tglatcmtaa
89496	atnltgatc(n)tanacttgg	atnltgat(c)taagacttgg
89499	ntgtatcmta(n)nacttggctg	ntgtatcmta(a)gacttggctg
89500	tglatcmtan(n)acttggctga	tglatcmta(g)acttggctga
90503	tttttatgt(n)gtctatttaa	tttttatgt(t)gtctatttaa
93328	tgtaataaa(n)caaaaangaa	tgtaataaa(a)caaaaangaa
93335	aaancaaaa(n)gaaaaaaa	aaancaaaa(a)gaaaaaaa
125580	atctgcttgg(n)taatttat	atctgcttgg(c)taatttat
135997	cttctccag(n)gtcagcaatg	cttctccag(t)gtcagcaatg
139861	gagccgagat(t)ccgccactgc	gagccgagat(c)ccgccactgc
151709	acatgaggt(n)gtcaatctt	acatgaggt(t)gtcaatctt
151725	agaaaggga(n)catlaaatc	agaaaggga(a)catlaaatc
199338	cctcccaacc(n)ntcccttgac	cctcccaacc(c)ntcccttgac
199339	ctcccaaccn(n)lcccttgacc	ctcccaacc(c)lcccttgacc
199375	ccctcccccct(n)ctatnctnc	ccctcccccct(c)ctatnctnc
199378	tcctccctnct(a)tnccntcccc	tcctccctcct(c)tcctccctcc
199380	ccctccnctat(n)cccctccct	ccctccnctat(c)cccctccct
199383	tcnctatnct(n)tcctccctcc	tcctctccct(c)tcctccctcc
200520	gtgcacctta(n)lctcatgtcc	gtgcacctta(a)lctcatgtcc
200566	aaatacagaa(n)caagagagcc	aaatacagaa(a)caagagagcc

----- Distribution of Quality < 40 Bases -----

#	bases	5	10	15	20	25	30	35	40
10001								*	*
9001								*	*
8001								*	*
7001								*	*
6001								*	*
5001								*	*
4001								*	*
3001								*	*
2001					*	*	*	*	*
1001				*	*	*	*	*	*
01		*	*	*	*	*	*	*	*

Phrap Value Range

Version: 1.01 qxf0.  
Location/Qualifiers  
1. 202103  
source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-711K1"  
complement(1..189)  
/rpt\_family="L1p"  
190..447  
/rpt\_family="HSMAR2"  
1208..1494  
/rpt\_family="AluJo"  
1944..2300  
/rpt\_family="MLT1B"  
3134..4152  
/rpt\_family="HSMAR2"  
4188..5034  
/rpt\_family="LTR12"  
5035..5135  
/rpt\_family="HSMAR2"  
5138..5296  
/rpt\_family="HERVL40"  
5857..5891  
/rpt\_family="(TTA)n"  
complement(5894..6183)  
/rpt\_family="AluJo"  
6547..6622  
/rpt\_family="HERVL40"  
6759..6947  
/rpt\_family="L1ME"  
7027..7173  
repeat\_region

Query Match 90.6%; Score 15.4; DB 87; Length 202103;  
Best Local Similarity 76.5%; Pred. No. 7.5e+02;  
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYGARGARATGAYCC 17  
11:11:11:11:11:11  
Db 1318 TTGAAGAGATGATCC 1302

RESULT 20  
AP001695/c 340000 bp DNA PRI 30-MAY-2000  
LOCUS AP001695 Homo sapiens genomic DNA, chromosome 21q, section 39/105.  
DEFINITION AP001695 AL163240 BA000005  
ACCESSION AP001695.1 GI:7768704  
VERSION  
KEYWORDS  
SOURCE Homo sapiens DNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T.,

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
2 (bases 1 to 340000)  
Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T.,  
Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., Soeda, E.,  
Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K.,  
Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R.,  
Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., and Schudy, A.  
The DNA sequence of human chromosome 21. The chromosome 21 mapping  
and sequencing consortium  
Nature 405 (6784), 311-319 (2000)  
20289799

TITLE  
JOURNAL  
Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The  
Chromosome 21 Mapping and Sequencing Consortium: \* RIKEN Genomic  
Sciences Center, Human Genome Research Group \* Institute of  
Molecular Biotechnology, Genome Analysis \* Keio University School  
of Medicine, Dept. of Molecular Biology \* GBF, Dept. of Genome  
Analysis \* Max-Planck Institute for Molecular Genetics (addresses  
see below)

COMMENT  
On May 30, 2000 this sequence version replaced gi:7717293.  
The chromosome 21 mapping and sequencing consortium consisting of  
\* RIKEN Genomic Sciences Center, Human Genome Research Group, \*  
Sagamihara 228-8555, Japan,  
\* e.mail: hattori@gsc.riken.go.jp  
\* URL: http://hgp.gsc.riken.go.jp/  
and  
\* Institute of Molecular Biotechnology, Genome Analysis, \*  
Beutenbergstrasse 11, D-07745 Jena, Germany,  
\* e.mail: gscj-submit@genome.imb-jena.de  
\* URL: http://genome.imb-jena.de/  
and  
\* Keio University School of Medicine, Molecular Biology, \* Tokyo  
160-8582, Japan,  
\* e.mail: nshimizuedmb-med.keio.ac.jp  
\* URL: http://www.dmb.med.keio.ac.jp/  
and  
\* GBF, Dept. of Genome Analysis,  
\* Mascheroder Weg 1, D-38124 Braunschweig, Germany, \* e.mail:  
info.genome@gbf.de  
\* URL: http://genome.gbf.de/  
and  
\* Max-Planck Institute for Molecular Genetics,  
\* Ihnestrasse 73, D-14195 Berlin, Germany,  
\* e.mail: info-chr21@molgen.mpg.de  
\* URL: http://chr21.rz-berlin.mpg.de/  
AL163240: Submitted (10-APR-2000).

FEATURES  
source  
1..340000  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
<1..46577  
/note="Accession No. AP001442"  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q21.2"  
/clone="p1715, 5' partial"  
/clone\_lib="CMP21 pl library"  
complement(1021..1140)  
/note="L1MA4"  
/rpt\_family="LINE/L1"

```
/rpt_type=DISPERSED
complement(1141. .1447)
/note="AluO"
repeat_region

/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(1448. .1605)
/note="L1MA4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
1971. .2404
/note="MT1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(2487. .2779)
/note="AluSc"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
3606. .3907
/note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
4155. .4318
/note="MT1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
4319. .4607
/note="AluSp"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
4608. .4835
/note="MT1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(5812. .6119)
/note="AluO"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(6634. .6734)
/note="MER5A"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
complement(8836. .9032)
/note="MER5A"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
complement(9523. .9820)
/note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(10813. .11086)
/note="L1MB3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(11095. .11195)
/note="U6"
/rpt_family="snRNA"
/rpt_type=DISPERSED
complement(11208. .11497)
/note="AluSp"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(11498. .11780)
/note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
13428. .13449
/note="(T)n"
/rpt_family="Simple-repeat"
/rpt_type=TANDEM
13483. .13746
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
```

```
STS
13676. .13790
/standard_name="SHGC-31514"
/note="AFMa356zh9"
Accession NO. M33112"
/db_xref="GDB:9732389"
complement(13747. .14027)
/note="AluDb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
14028. .14430
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
14431. .14742
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
14743. .14806
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
15768. .16072
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
17881. .18012
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(18041. .18207)
/note="MER91A"
/rpt_family="DNA/MER1_type?"
/rpt_type=DISPERSED
complement(18416. .18684)
/note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(19782. .19899)
/note="L1MC5"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
20230. .20272
/note="(CA)n"
repeat_region
```

Query Match 90.6%; Score 15.4; DB 90; Length 340000;  
Best Local Similarity 76.5%; Pred. No. 7.8e+02;  
Matches 13; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYGARGARATGGAYCC 17  
DB 273372 TTTGAGGAATCGATCC 273356

RESULT 21  
LOCUS G01707 355 bp DNA STS 07-DEC-1994  
DEFINITION chicken STS ADL288.  
ACCESSION G01707  
VERSION G01707.1 GI:595231  
KEYWORDS STS sequence; primer; sequence tagged site.  
SOURCE chicken vector-pBluescript II KS+ host=E. coli.  
ORGANISM Gallus gallus  
Eukaryotae; Mitochondria; Metazoa; Chordata; Vertebrata; Tetrapoda;  
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 355)  
AUTHORS Cheng, H.H.  
JOURNAL Unpublished (1994)  
COMMENT Synonyms: B407  
Contact: Hans H. Cheng  
Avian Disease and Oncology Laboratory  
USDA-ARS  
3606 E. Mount Hope Rd, East Lansing, MI 48823, USA  
Tel: 5173376758



Fax: 5173376776  
Email: hcheng@pilot.msu.edu

Primer A: AAACCTGCTCCCAACAT  
Primer B: ATCCCCACGTAATCTAT  
STS size: 152  
PCR profile:

Denaturation: 94 degrees C for 1 minute  
Annealing: 48 degrees C for 1 minute  
Polymerization: 72 degrees C for 1 minute  
PCR Cycles: 30  
Thermal Cycler: MJ Research

Protocol:  
Template: 50-100 ng  
Primer: 0.1 uM  
dNTPs: 200 uM  
Taq polymerase: 0.4 units/ul  
Total Vol: 25 ul

## Buffer:

MgCl2: 1.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 9.0  
Triton X-100: 0.1 %

FEATURES  
source location/Qualifiers  
1..355

STS  
primer\_bind /organism="Gallus gallus"  
202..353

primer\_bind 202..221  
complement(334..353)

BASE COUNT 75 a 69 c 86 g 124 t 1 others  
ORIGIN

Query Match 84.7%; Score 14.4; DB 54; Length 355;  
Best Local Similarity 75.0%; Pred. No. 1.6e+03;  
Matches 12; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAYC 16  
||:|:|:|:|:|:|:|:|:|  
Db 301 TTTGAGGAATGATC 316

## RESULT 22

MMHBEGFL4

LOCUS 710 bp DNA ROD 19-JUN-1996

DEFINITION Mus musculus heparin-binding epidermal growth factor-like growth factor gene, exon 5 and 6, complete cds.

ACCESSION U39192  
VERSION U39192.1 GI:1236398KEYWORDS  
SEGMENT 4 of 4  
SOURCE house mouse strain=129SVJ.

ORGANISM

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 710)  
Harding, P.A., Brigstock, D.R., Shen, L., Crissman-Combs, M.A. and Besner, G.E.TITLE  
AUTHORS Characterization of the gene encoding murine heparin-binding epidermal growth factor-like growth factorJOURNAL  
MEDLINE  
REFERENCE Gene 169 (2), 291-292 (1996)  
96194822AUTHORS 2 (bases 1 to 710)  
Harding, P.A., Brigstock, D.R., Shen, L., Crissman-Combs, M.A. and Besner, G.E.TITLE  
JOURNAL Direct SubmissionFEATURES  
SOURCE Submitted (23-OCT-1995) Paul A. Harding, Pediatric Surgery, Childrens Hospital, 700 Childrens Drive, Columbus, OH 43205, USA  
location/Qualifiers  
1..710  
/organism="Mus musculus"

/strain="129SVJ"  
/db\_xref="taxon:10090"  
/chromosome="18"  
/note="129SVJ mouse genomic library in the lambda FIX II vector"  
join(U39189.1:163..208,U39189.1:412..585,U39190.1:68..245,U39191.1:155..310,135..207)  
/codon\_start=1  
/product="heparin-binding epidermal growth factor-like growth factor"  
/protein\_id="AAC52617.1"  
/db\_xref="GI:1236400"

/translation="MKLPSVWMLKFLAVALSALVTGESLERLRGLAATSNDDPT  
GSTNQLPTGGDRAQGVQDLEGTDLNLFKAFSSKPGGLATPSKERNGKKKGGGLG  
KKRDPCLRKKYKDYCIHGECRYLQEFRTSPCKCLPGYHGRCHGLTPVENPLYTYDHT  
TVLAVAVVLSVCLVIYVGLMFRYHRRGYDLESEKVKLGVASSH"  
<1..134  
/note="intron 4 is approximately 825 bp; heparin-binding epidermal growth factor-like growth factor"  
/number=4  
135..225  
/number=5  
/product="heparin-binding epidermal growth factor-like growth factor"  
225..497  
/note="heparin-binding epidermal growth factor-like growth factor"  
/number=5  
498..>710  
/number=6  
/product="heparin-binding epidermal growth factor-like growth factor"

intron  
exon  
exon  
intron

BASE COUNT 170 a 168 c 187 g 185 t  
ORIGIN

Query Match 84.7%; Score 14.4; DB 94; Length 710;  
Best Local Similarity 75.0%; Pred. No. 1.7e+03;  
Matches 12; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTYGARGARATGAYC 16  
||:|:|:|:|:|:|:|:|:|  
Db 40 TTTGAGGAATGATC 55

## RESULT 23

CNS01GZU

LOCUS 963 bp DNA STS 17-FEB-2000

DEFINITION Anopheles gambiae STS T7 end of clone 07M16 of NotreDame1 library from strain PEST of Anopheles gambiae (African malaria mosquito), sequence tagged site.

ACCESSION AL143803  
VERSION AL143803.1 GI:7001965KEYWORDS  
SOURCE STS.  
ORGANISM African malaria mosquito.REFERENCE  
AUTHORS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae; Anopheles.  
1 (bases 1 to 963)  
Genoscope.TITLE  
AUTHORS Direct Submission  
JOURNAL Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.frREFERENCE  
AUTHORS - Web : www.genoscope.cns.fr)  
2 (bases 1 to 963)  
Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissenbach, J.TITLE  
JOURNAL Direct Submission  
COMMENT Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France  
This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut



BASE COUNT                    /clone\_lib="AZ0AA"  
 ORIGIN                    262 a    245 c    196 g    380 t    1 others

Query Match                    84.7%;    Score 14.4;    DB 53;    Length 1084;  
 Best Local Similarity    75.0%;    Pred. No. 1.8e+03;  
 Matches    12;    Conservative    4;    Mismatches    0;    Indels    0;    Gaps    0;

QY    1    TTYGARGARATGAYC 16  
 ||:||:||:||:||:||:||||  
 Db    426    TTCGAAGAAATGATC 411

Search completed: May 15, 2001, 11:13:37  
 Job time: 3608 sec